Air-BNB
Travel Data Analysis

~Vineet Patel

Objective:

Air BNB has aimed to expand on travelling possibilities and present more unique, personalized ways of experiencing destinations. The analysis aims to draw insights from the data obtained from the listing activity of homestays in San Diego, California for 2019. Some research questions need to be answered with respect to all the listings of past booking information.

Problem Statement:

Since 2008, guests and hosts have used AirBNB to expand on travelling possibilities and present more unique, personalized way of experiencing the world. This dataset describes the listing activity and metrics in San Diego, California for 2019.

➢ **Data Extraction**

- The data extracted from csv file and loaded into a Pandas DataFrame.
- Initial data exploration includes checking the dataset's shape and inspecting the first few rows.

➢ **Data Transformation**

- Drop columns with most missing values (country, borough, bathrooms and minstay)
- converted last modified column from object to datetime and Extracted Date, Month and Year as new columns.
- Calculated Total earnings, Monthly earnings.
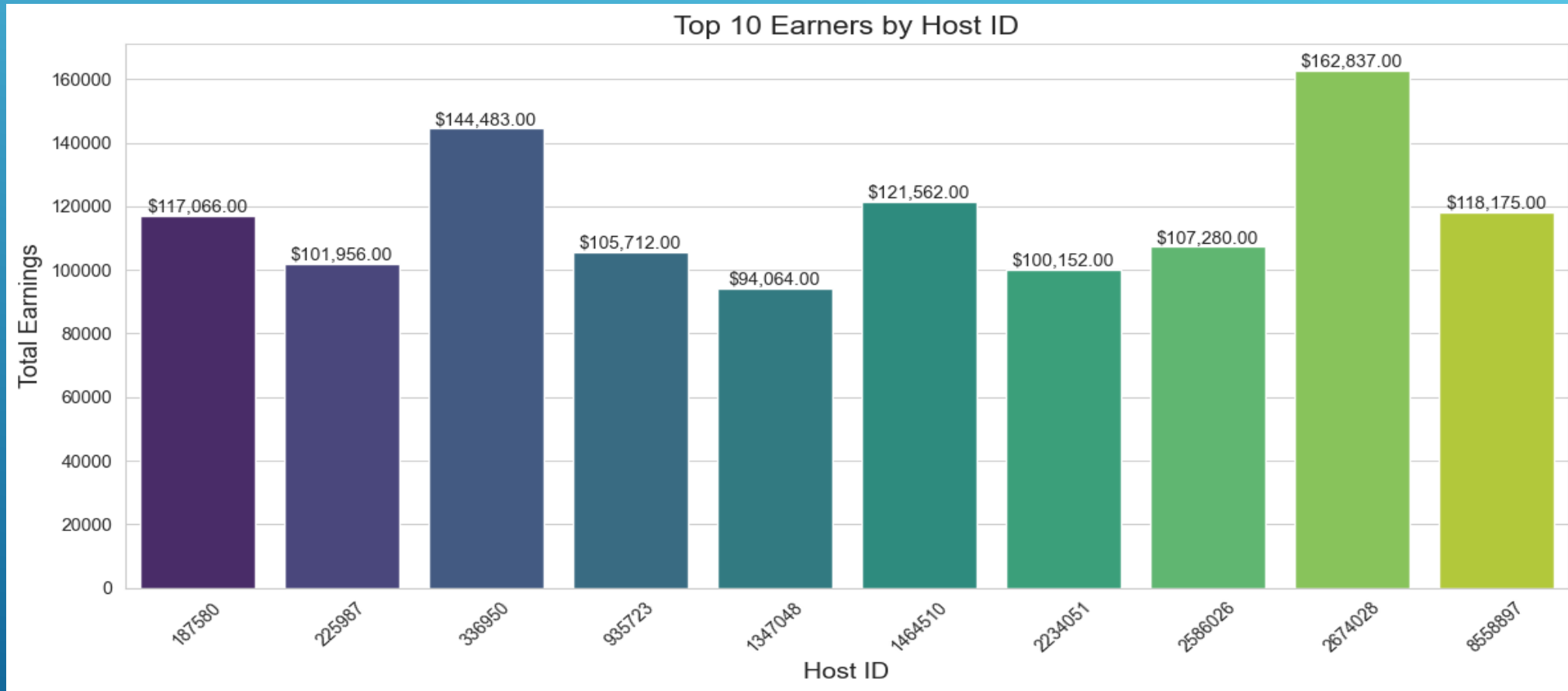- Stored the cleaned Transformed data to new csv file.

➢ **Load Cleaned Dataset:**

Exploratory Data Analysis (EDA)

1. **Regarding the Host:**
   A. Who are top earners?

**Host IDs 2674028, 336950, and 1464510 are the 3 highest earners.**

B. Is there any relationship between monthly earning and price

- **Correlation between Monthly Earning and Price: 0.15, P-value: 0.00**
- **This result suggests that while there is a statistically significant relationship between prices and monthly earnings, the strength of the relationship is weak.**
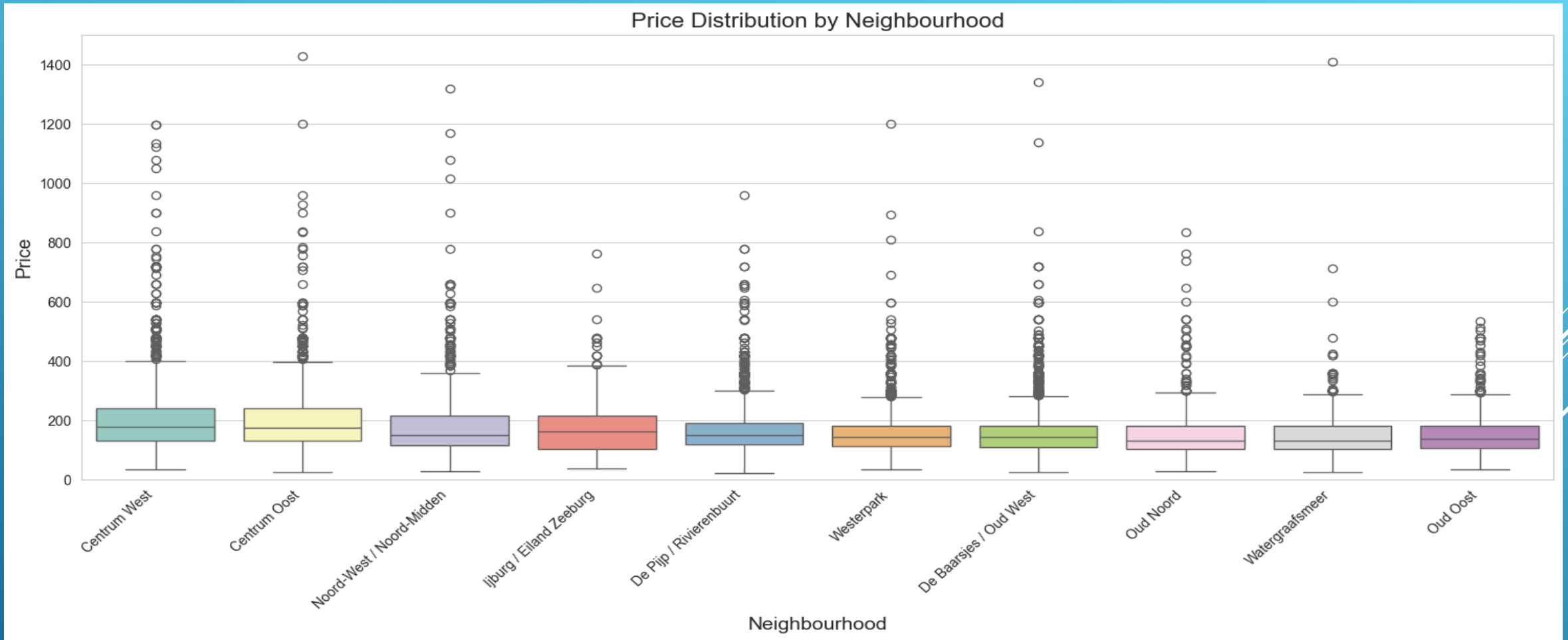
2. Regarding the Neighbourhood:

A.   Any particular location getting maximum number of bookings?

Top 10 Neighborhoods getting Maximum Number of Bookings
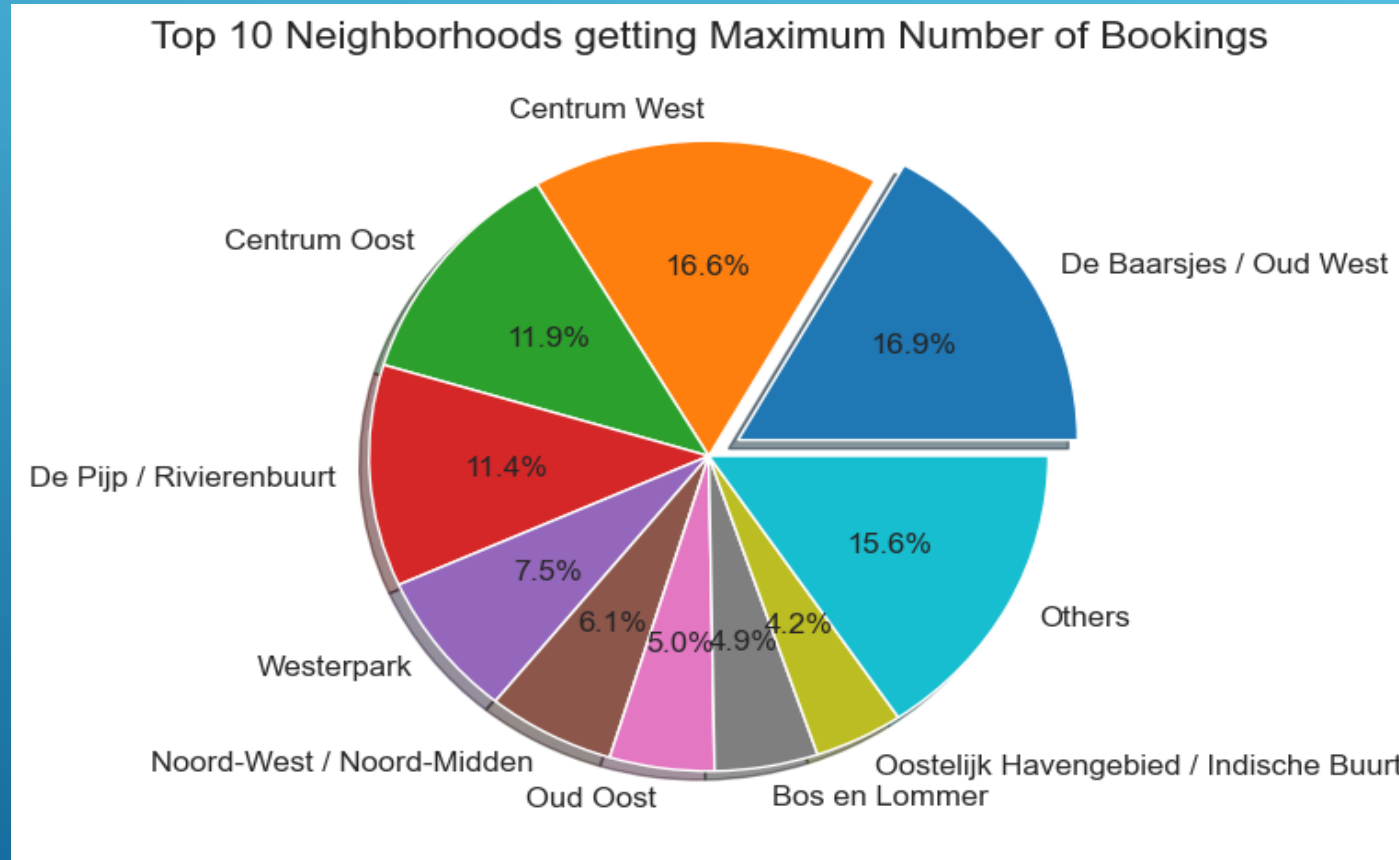
The "De Baarsjes / Oud West" neighbourhood is receiving the highest number of bookings, accounting for 16.9% of overall bookings.

B. Price distribution with respect to location

Price Distribution by Neighbourhood

There is a statistically significant difference in average prices between "De Baarsjes / Oud West" and "De Pijp / Rivierenbuurt," with prices in "De Baarsjes / Oud West" being lower (t-statistic: -4.208, p-value: 2.62e-05).

3. Regarding the reviews

A. Relationship between Quality and Price

Price vs Overall Satisfaction Scatter Plotwith Regression Line
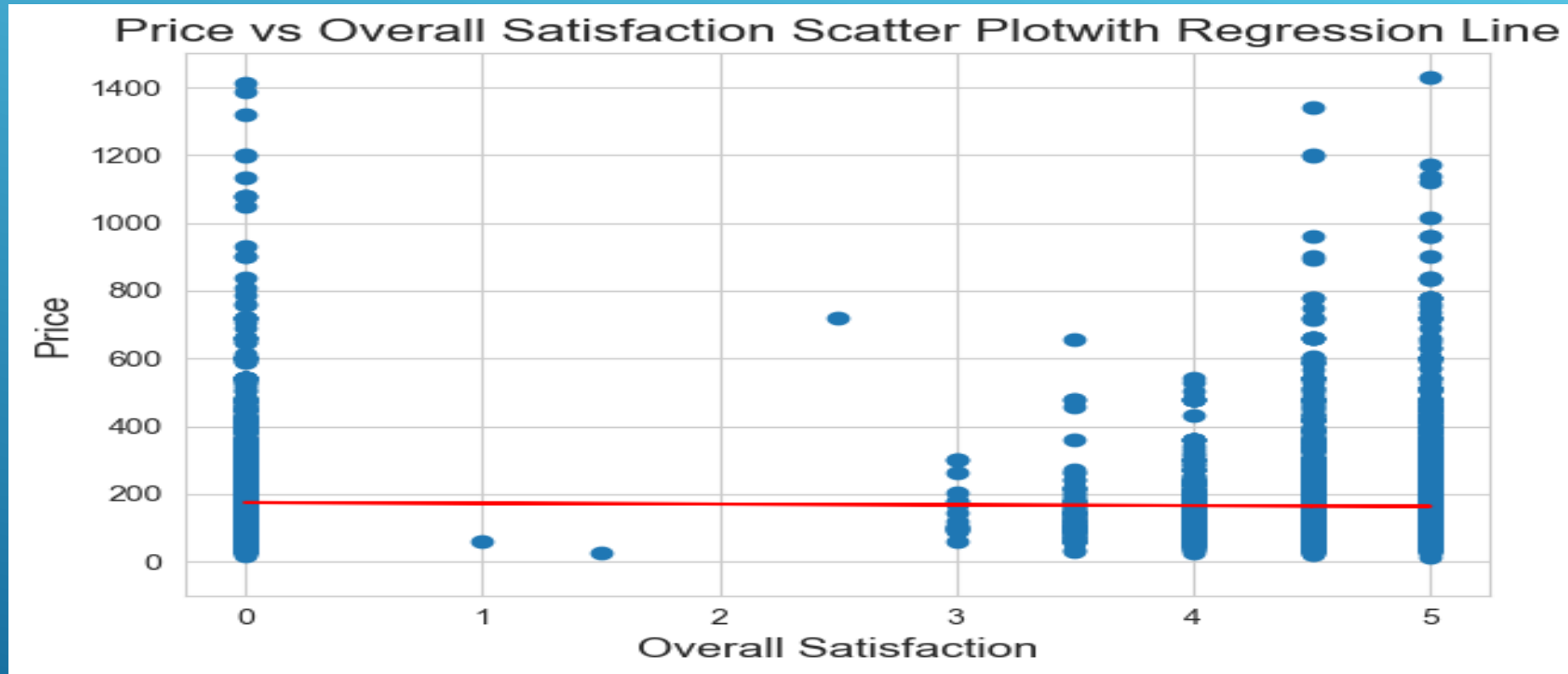
Correlation Coefficient (-0.045): This indicates a very weak negative correlation between overall satisfaction scores and price. This means that as prices increase, satisfaction scores tend to decrease slightly, but the relationship is very weak.
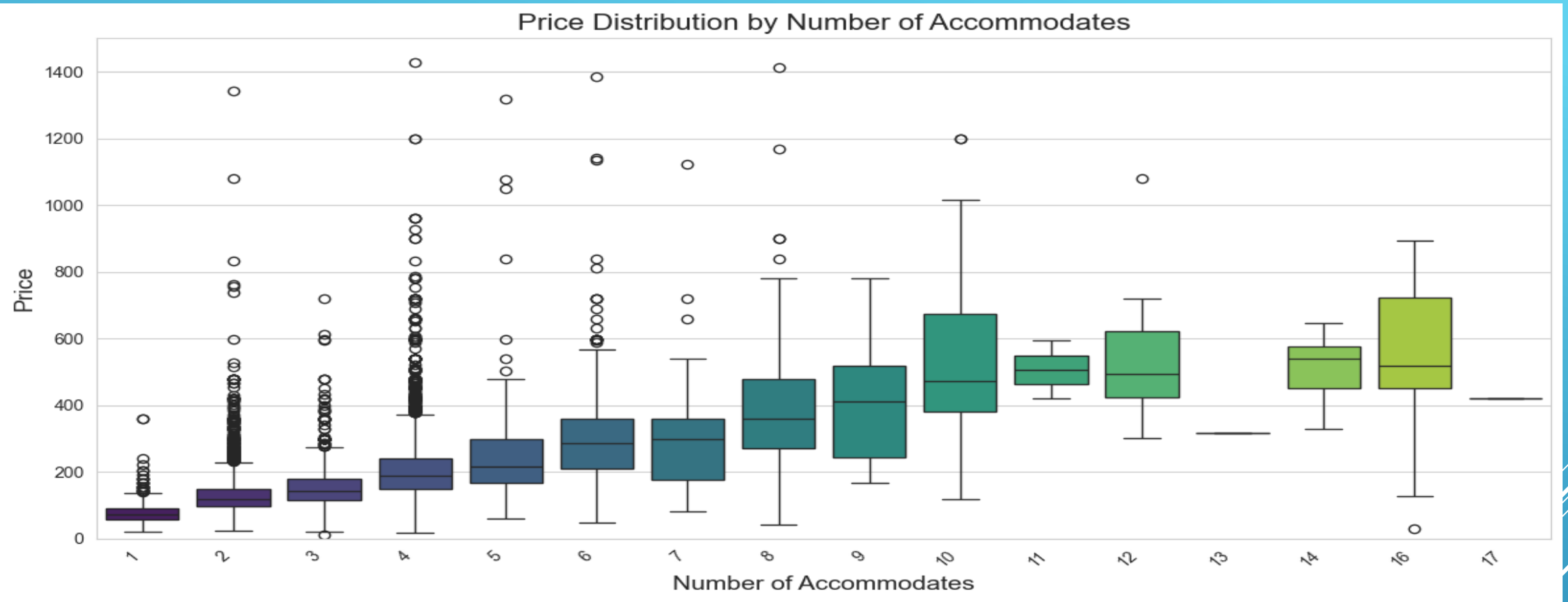
# 4. Price vs Amenities:

A. Average Price vs Room type
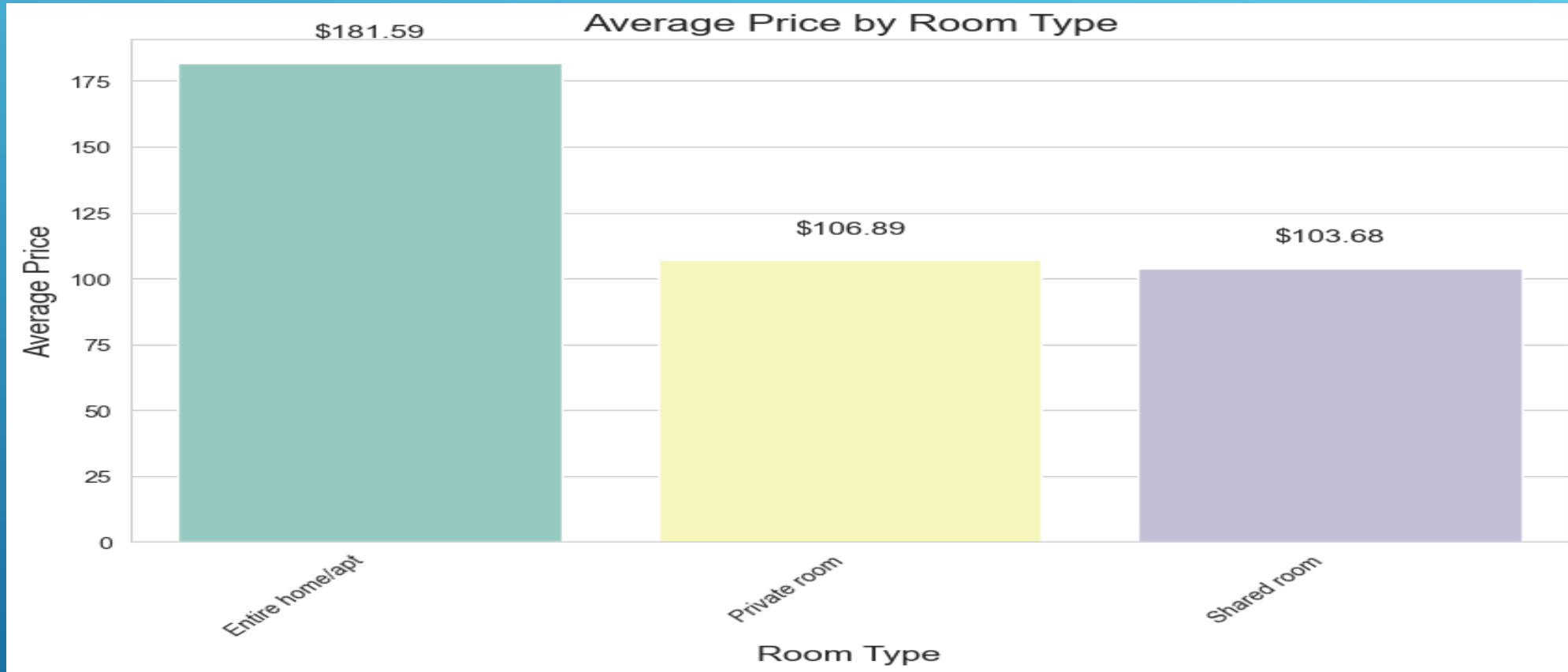
- There is significant difference in prices between Entire home/apt and private rooms and also between Entire home/apt and shared rooms.
- There is No significant difference in prices between shared rooms and private rooms.

B. Average Price vs Accommodates

Price Distribution by Number of Accommodates

▶ A correlation coefficient of 0.5 between price and accommodates indicates a moderate positive relationship. This means that as the number of people a property can accommodate increases, the price tends to increase as well, though it's not a perfect relationship.

▶

The comparisons indicate that the price differences for most of the "accommodates" values are statistically significant. There are a few comparisons with no significant differences, which may indicate similar pricing for those specific groups.

Price Distribution by Number of Bedrooms
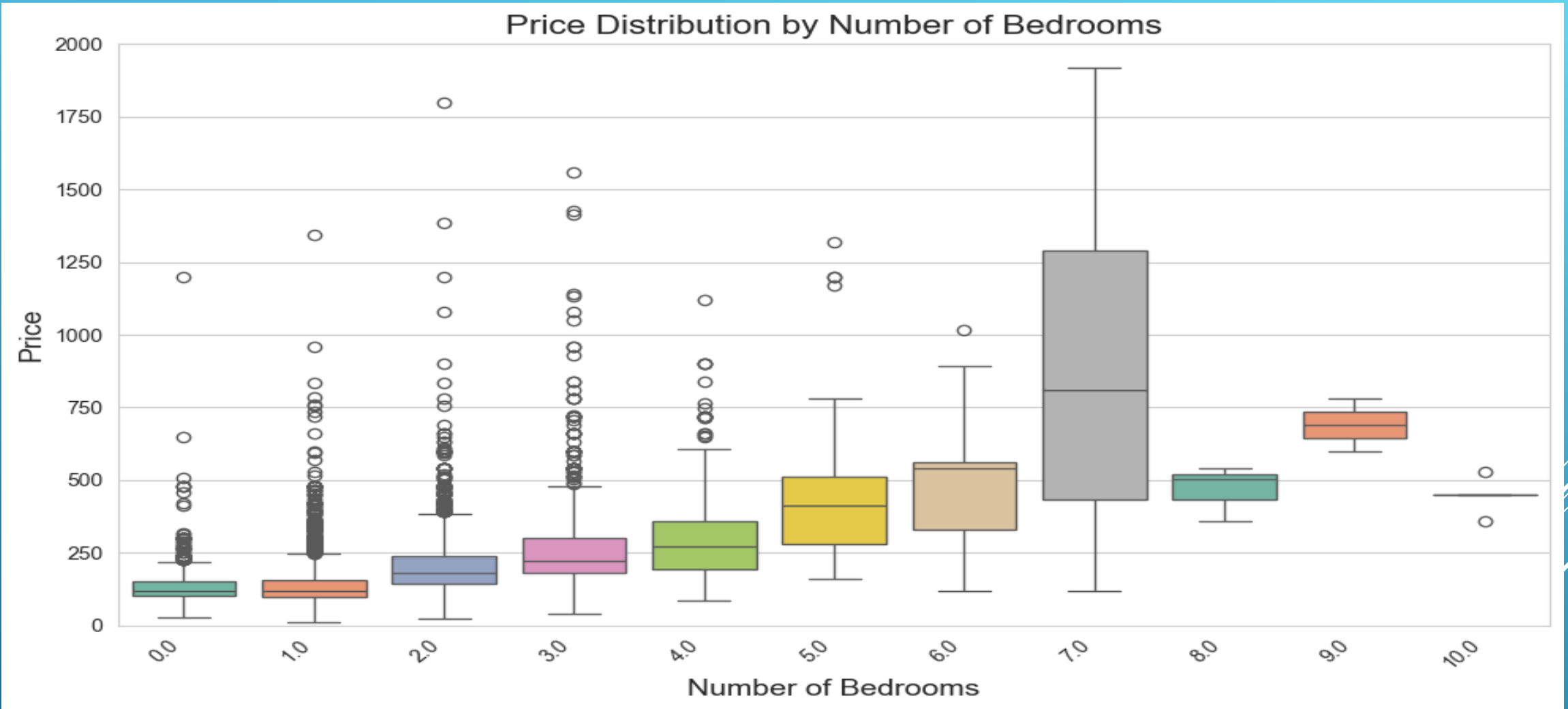
A correlation coefficient of 0.45 between price and bedrooms suggests a moderate positive relationship. This indicates that, generally, as the number of bedrooms increases, the price also tends to increase, but the correlation isn't extremely strong.

Conclusion:

1. Host Earnings:

A. Top Earners: The highest earners are Host IDs 2674028, 336950, and 1464510. These hosts likely have multiple listings, high-priced listings, or both, potentially offering premium services or being located in high-demand areas.

B. Monthly Earnings vs. Prices: There is a weak positive correlation (0.153) between monthly earnings and prices, meaning higher prices tend to lead to higher earnings, but the relationship is not very strong. Despite this, the correlation is statistically significant (p-value of 2.197e-84), indicating a genuine relationship.

2. Neighbourhoods:

A. Bookings: The "De Baarsjes / Oud West" neighbourhood leads in bookings with 16.9% of the total bookings, closely followed by "Centrum West" with 16.6%. These areas are likely popular due to their amenities, attractions, or high number of listings.

B. Price Differences by Neighbourhood: There is a statistically significant difference in average prices between "De Baarsjes / Oud West" and "De Pijp / Rivierenbuurt," with prices in "De Baarsjes / Oud West" being lower (t-statistic: -4.208, p-value: 2.62e-05).

3. Reviews:

A. Quality vs. Price: There is a very weak negative correlation (-0.045) between overall satisfaction scores and price, suggesting that higher prices are associated with slightly lower satisfaction scores. This relationship is statistically significant (p-value: 7.32e-10).

4. Price Analysis

A. Room Types:

a. Shared Room vs. Entire Home/Apt: Shared rooms are significantly cheaper than entire homes/apartments (t-statistic: -5.3936, p-value: 7.0101e-08).

b. Shared Room vs. Private Room: No significant price difference between shared rooms and private rooms (t-statistic: -0.5092, p-value: 0.6106).

c. Entire Home/Apt vs. Private Room: Entire homes/apartments are significantly more expensive than private rooms (t-statistic: 38.6702, p-value: 0.0).

B. Price vs. Accommodates: There is a moderate positive correlation (0.5) between price and the number of people a property can accommodate, indicating that as accommodation capacity increases, so does the price. Differences in pricing based on accommodation capacity are generally statistically significant.

C. Price vs. Bedrooms: A moderate positive correlation (0.45) exists between price and the number of bedrooms, meaning that more bedrooms typically lead to higher prices. Most comparisons of prices based on bedroom count are statistically significant.

D. Price vs. Location: Price variations by location have been previously covered in the neighbourhood analysis.

This summary provides an overview of key insights related to host earnings, neighbourhood popularity, review ratings, and pricing trends based on various factors.

This findings can help hosts optimize their listings and pricing strategies based on location, amenities, and guest preferences. They provide valuable insights for potential hosts on where to list their properties and how to enhance guest satisfaction.